

Word, Sense and Contextualized Embeddings: Vector Representations of Meaning in NLP

Jose Camacho-Collados



Cardiff University, 18 March 2019

Outline

❖ Background

- *Vector Space Models (word embeddings)*
- *Lexical resources*

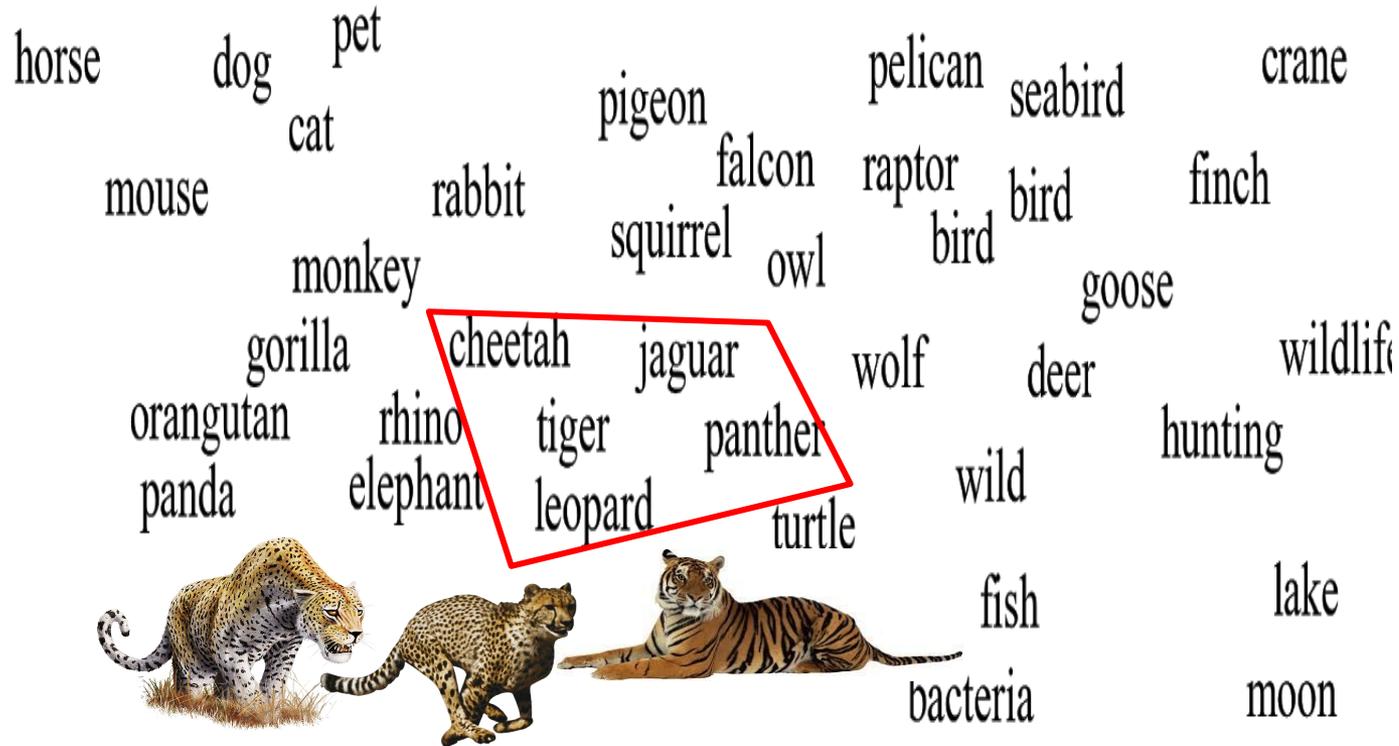
❖ Sense representations

- *Knowledge-based: NASARI, SW2V*
- *Contextualized: ELMo, BERT*

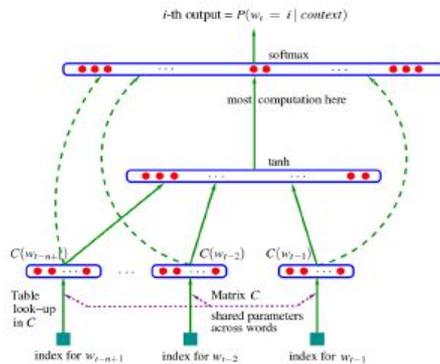
❖ Applications

Word vector space models

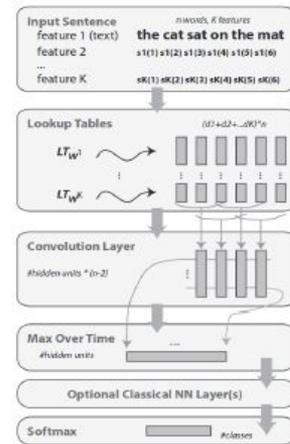
Words are represented as vectors: semantically similar words are close in the vector space



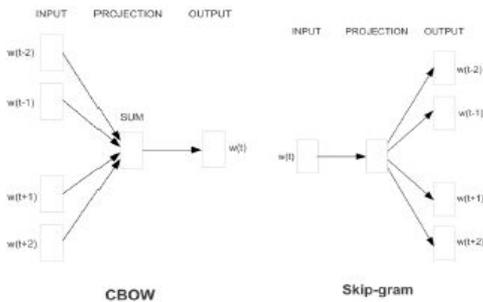
Neural networks for learning word vector representations from text corpora -> word embeddings



Bengio et al. (2003)



Collobert & Weston (2008)



Mikolov et al. (2013)

Probability and Ratio	$k = \text{solid}$	$k = \text{gas}$	$k = \text{water}$
$P(k \text{ice})$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k \text{steam})$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k \text{ice})/P(k \text{steam})$	8.9	8.5×10^{-2}	1.36

Pennington et al. (2014)

Why word embeddings?

Embedded vector representations:

- are compact and fast to compute
- preserve important relational information between words (actually, meanings):

$$\textit{king} - \textit{man} + \textit{woman} \approx \textit{queen}$$

- are geared towards general use

Applications for word representations

- Syntactic parsing (Weiss et al. 2015)
- Named Entity Recognition (Guo et al. 2014)
- Question Answering (Bordes et al. 2014)
- Machine Translation (Zou et al. 2013)
- Sentiment Analysis (Socher et al. 2013)

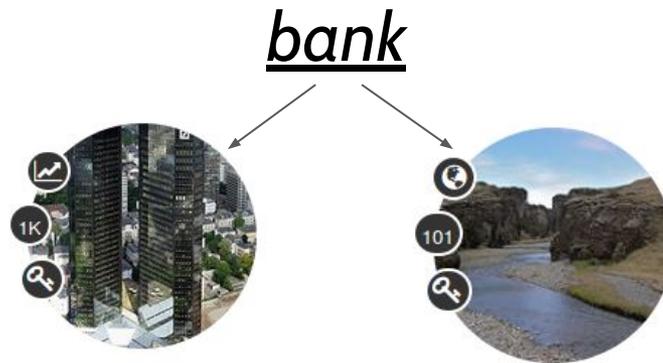
... and many more!

AI goal: language understanding



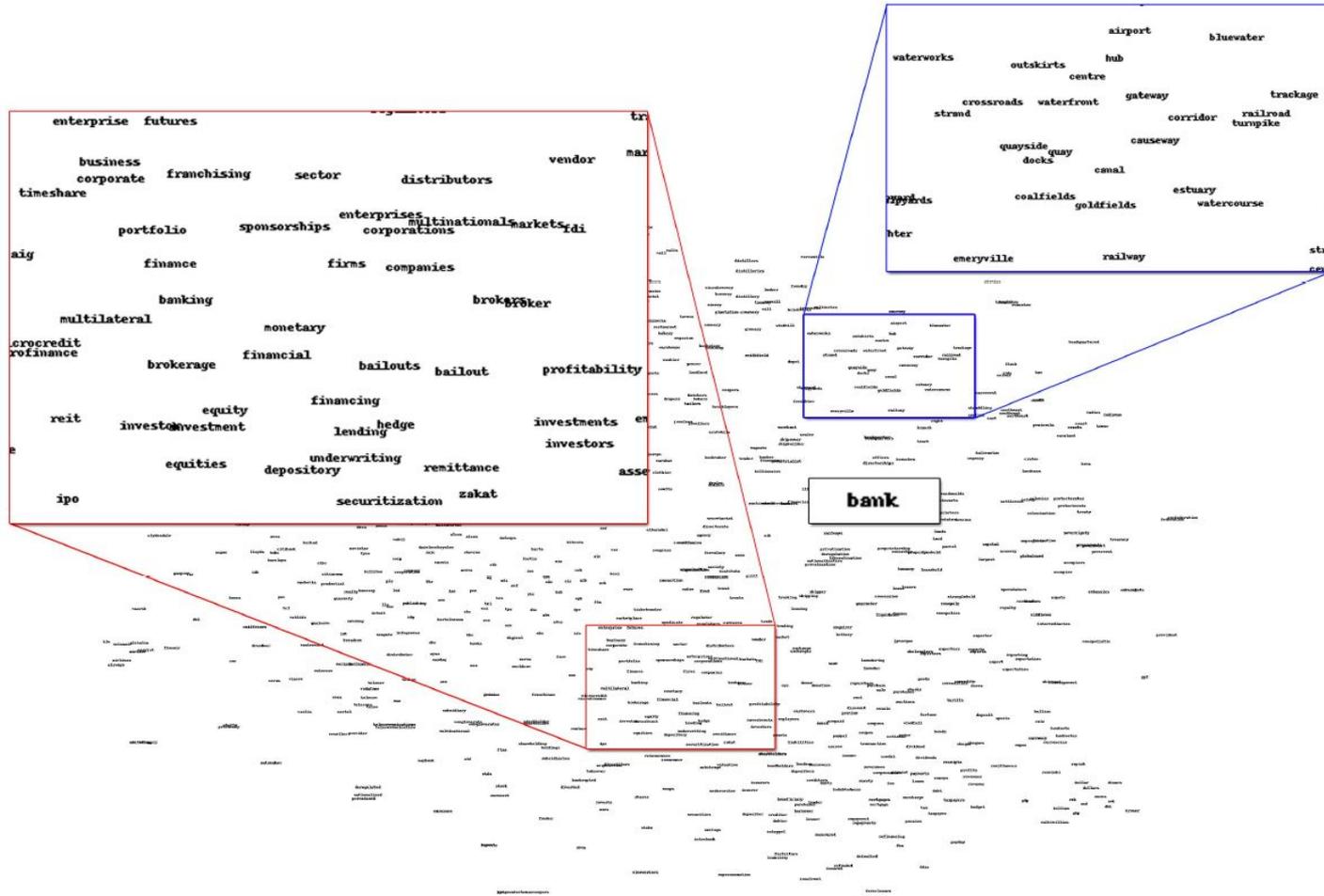
Limitations of word embeddings

- Word representations cannot capture ambiguity. For instance,

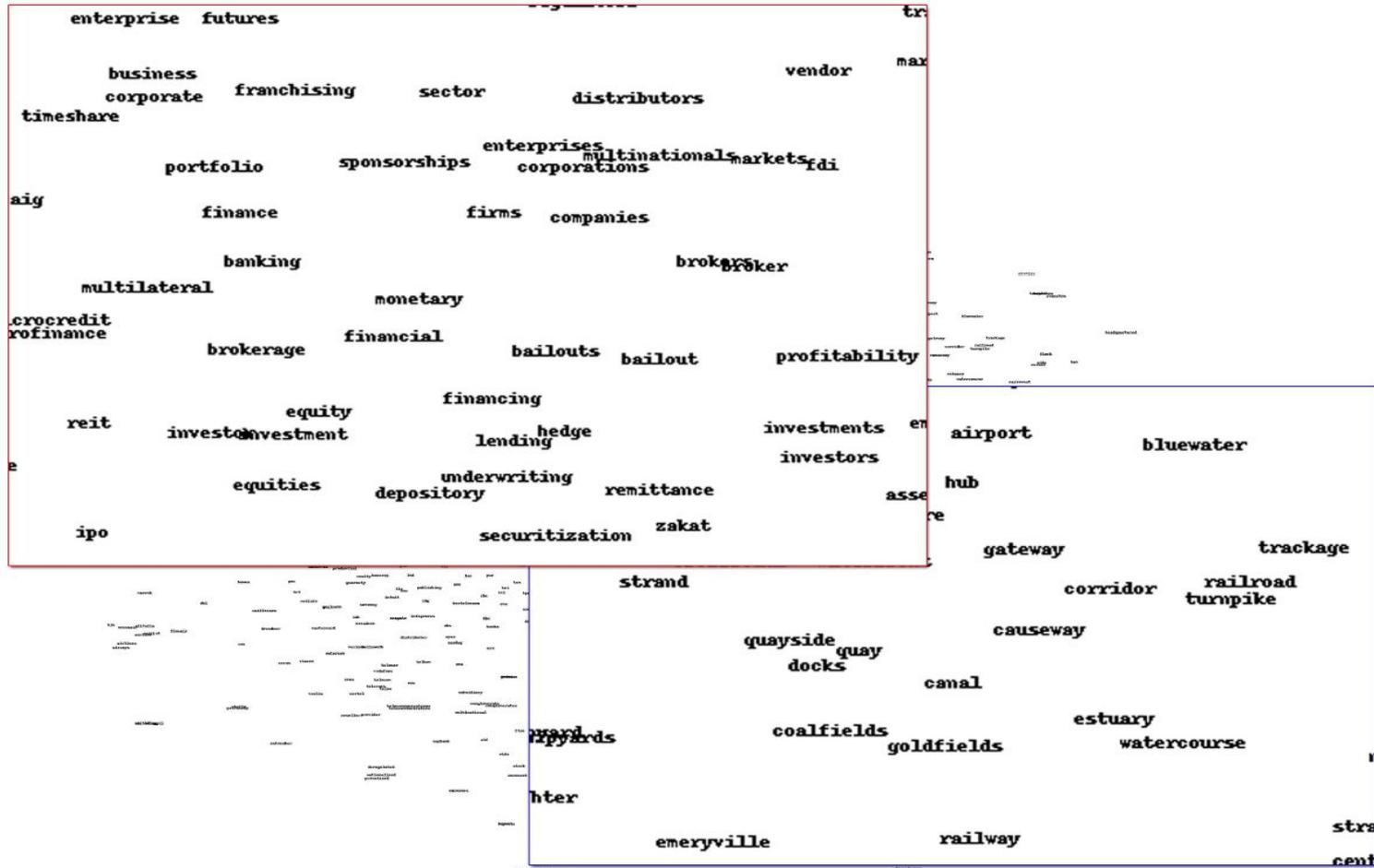


Problem 1:

word representations cannot capture ambiguity



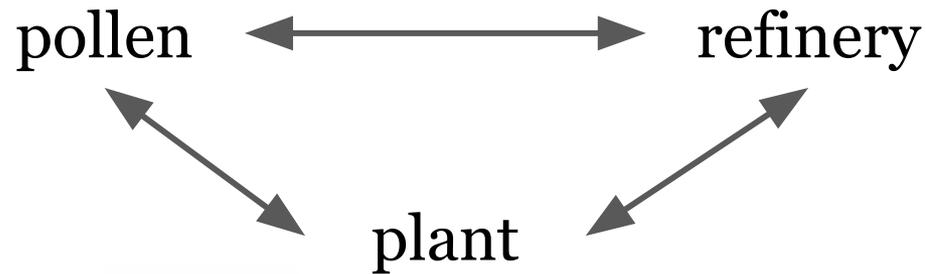
Problem 1: word representations cannot capture ambiguity



Word representations and the triangular inequality

Example from Neelakantan et al (2014)

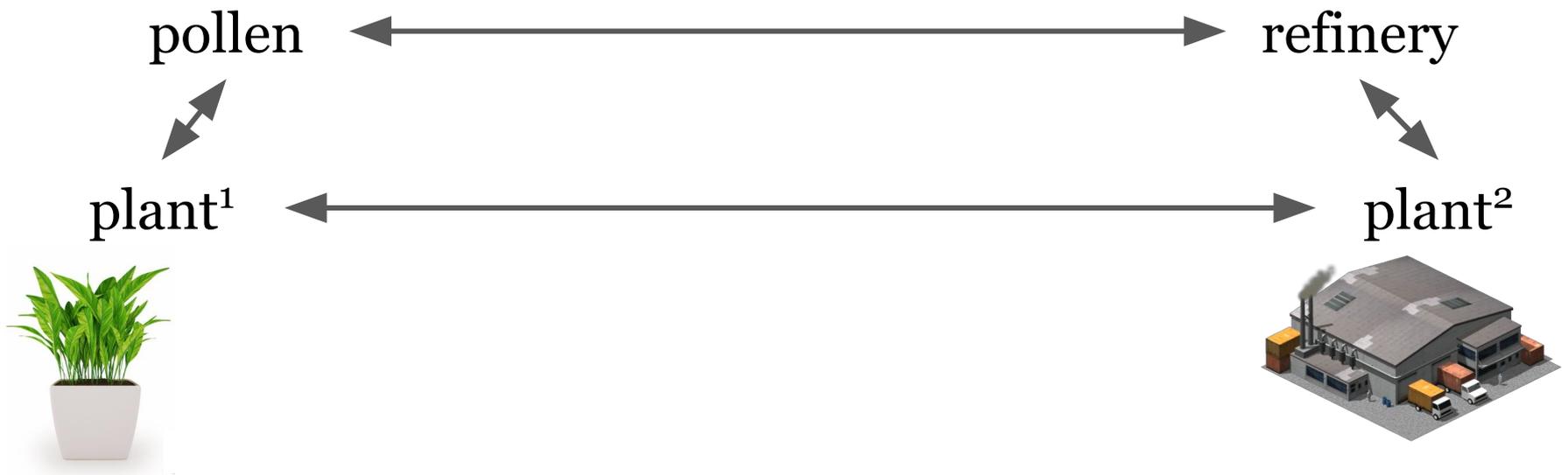
For distance d , $d(a, c) \leq d(a, b) + d(b, c)$.



Word representations and the triangular inequality

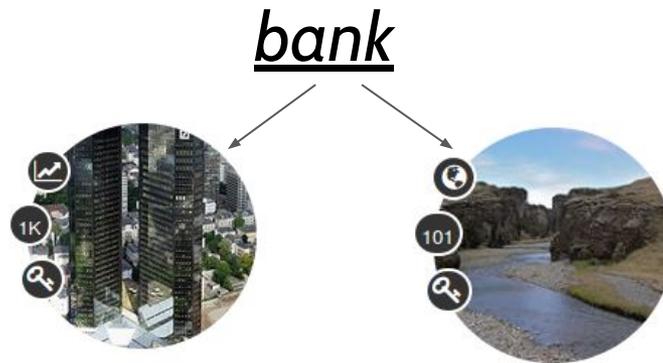
Example from Neelakantan et al (2014)

For distance d , $d(a, c) \leq d(a, b) + d(b, c)$.



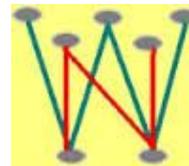
Limitations of word representations

- They cannot capture ambiguity. For instance,



-> They neglect rare senses and infrequent words

- Word representations do not exploit knowledge from existing lexical resources.



Motivation: Model senses instead of only words

*He withdrew money from the **bank**.*



Motivation: Model senses instead of only words

*He withdrew money from the **bank**.*



bank#1



bank#2



Motivation: Model senses instead of only words

*He withdrew money from the **bank**.*



bank#1



bank#2

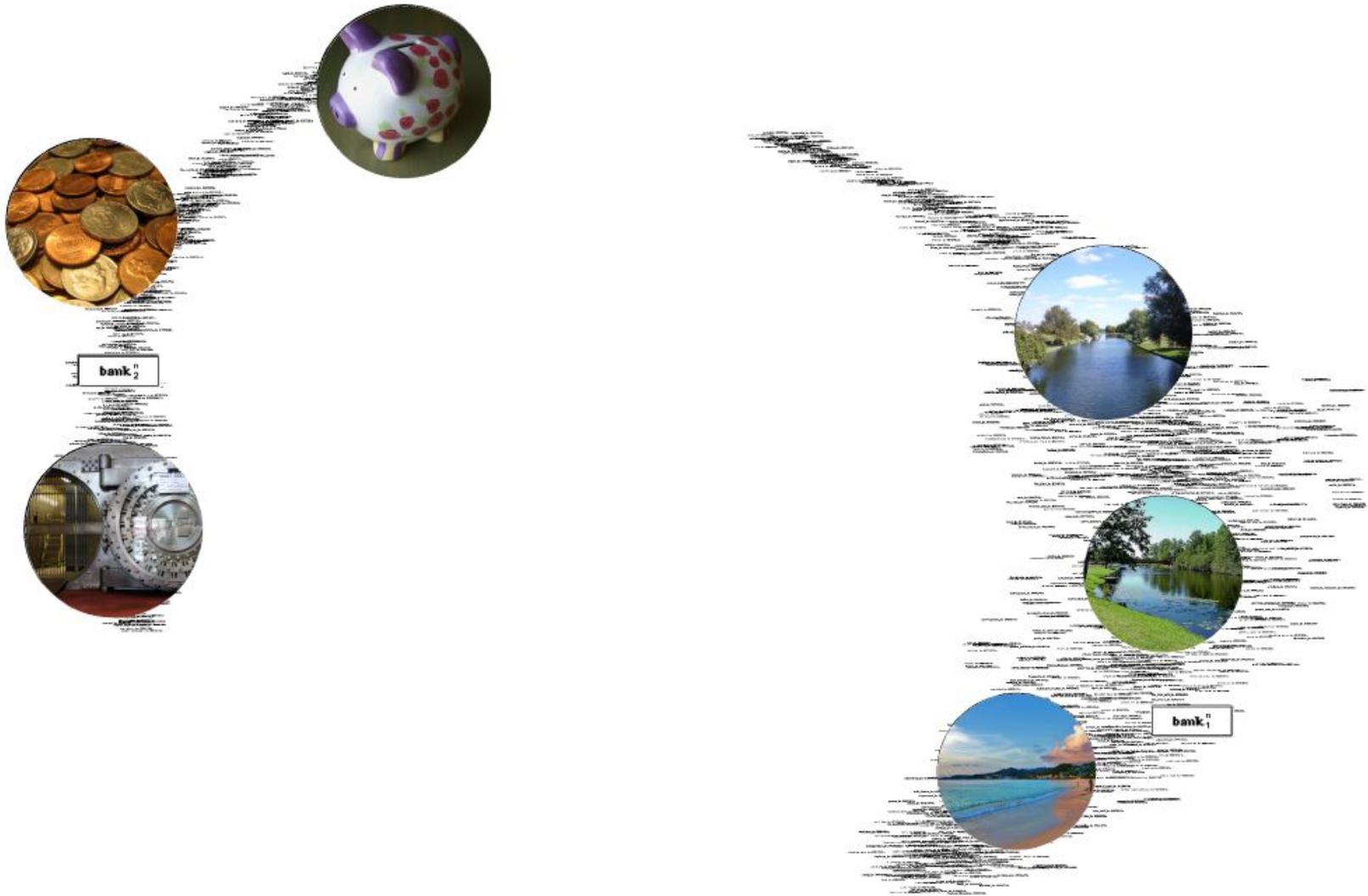




a Novel Approach to a Semantically-Aware Representations of Items

<http://lcl.uniroma1.it/nasari/>

Key goal: obtain sense representations



Key goal: obtain sense representations

- Nome
- Verbo

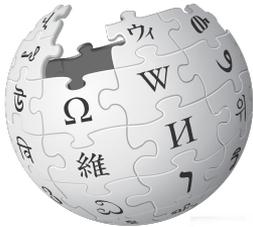
Nome

	bank, streambank Sloping land (especially the slope beside a body of water) ID: 00008363n Concetto	AR ضفة, حافة ZH 岸, 河边 FR berge, rive IT riva, argine, sponda
	bank, depository financial institution, banking company A financial institution that accepts deposits and channels the money into lending activities ID: 00008364n Concetto	AR مصرف (أموال), بنك, البنك ZH 銀行, 银行, 存放款金融机构 FR banque, institution financière de dépôt, établissement bancaire IT banca, banco, cassa
	bank A long ridge or pile ID: 00008365n Concetto	FR banc IT banco
	bank An arrangement of similar objects in a row ID: 00008366n Concetto	
	bank A supply or stock held in reserve for future use (especially in emergencies) ID: 00008367n Concetto	ZH 储备金 FR banque IT banca

We want to create a separate representation for each entry of a given word

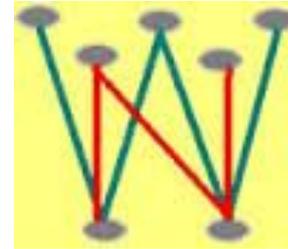
Idea

Encyclopedic knowledge



WIKIPEDIA
The Free Encyclopedia

Lexicographic knowledge



WordNet



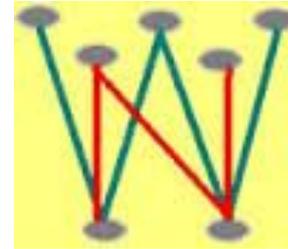
Idea

Encyclopedic knowledge



WIKIPEDIA
The Free Encyclopedia

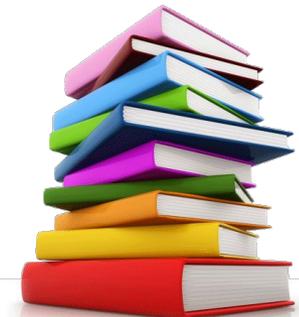
Lexicographic knowledge



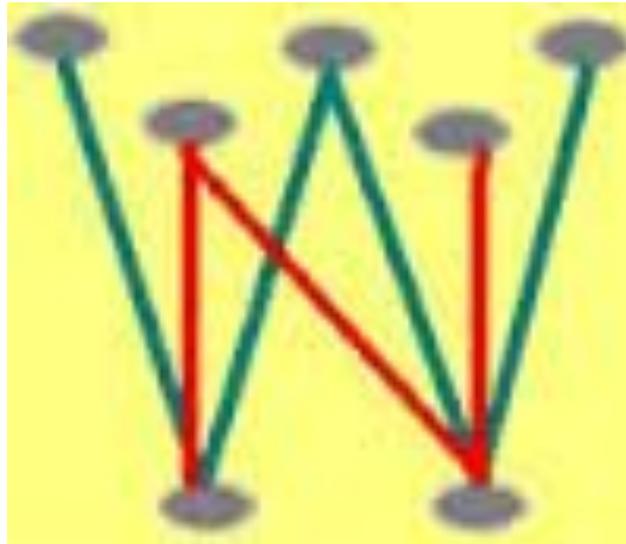
WordNet



Information from text corpora

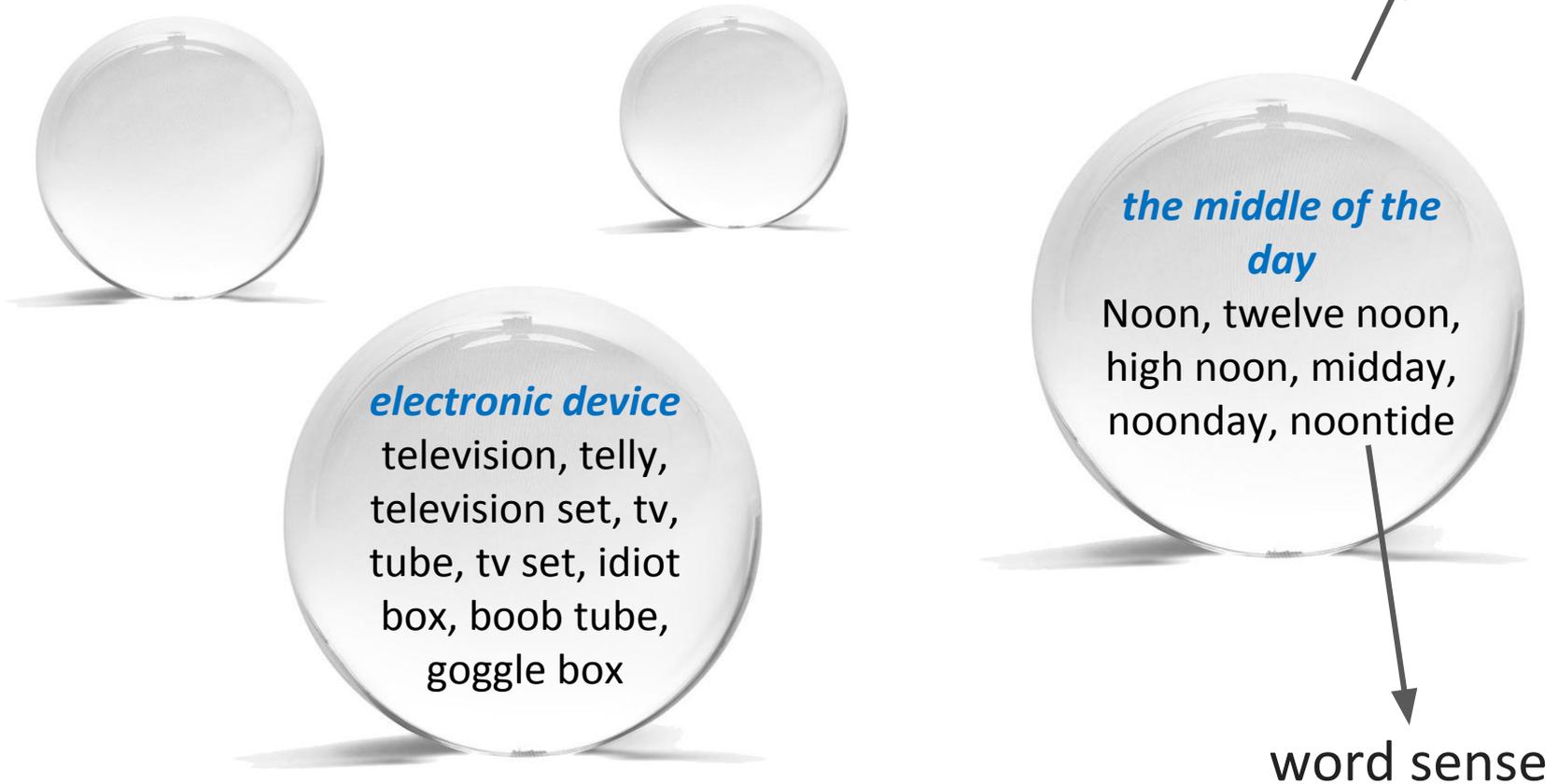


WordNet

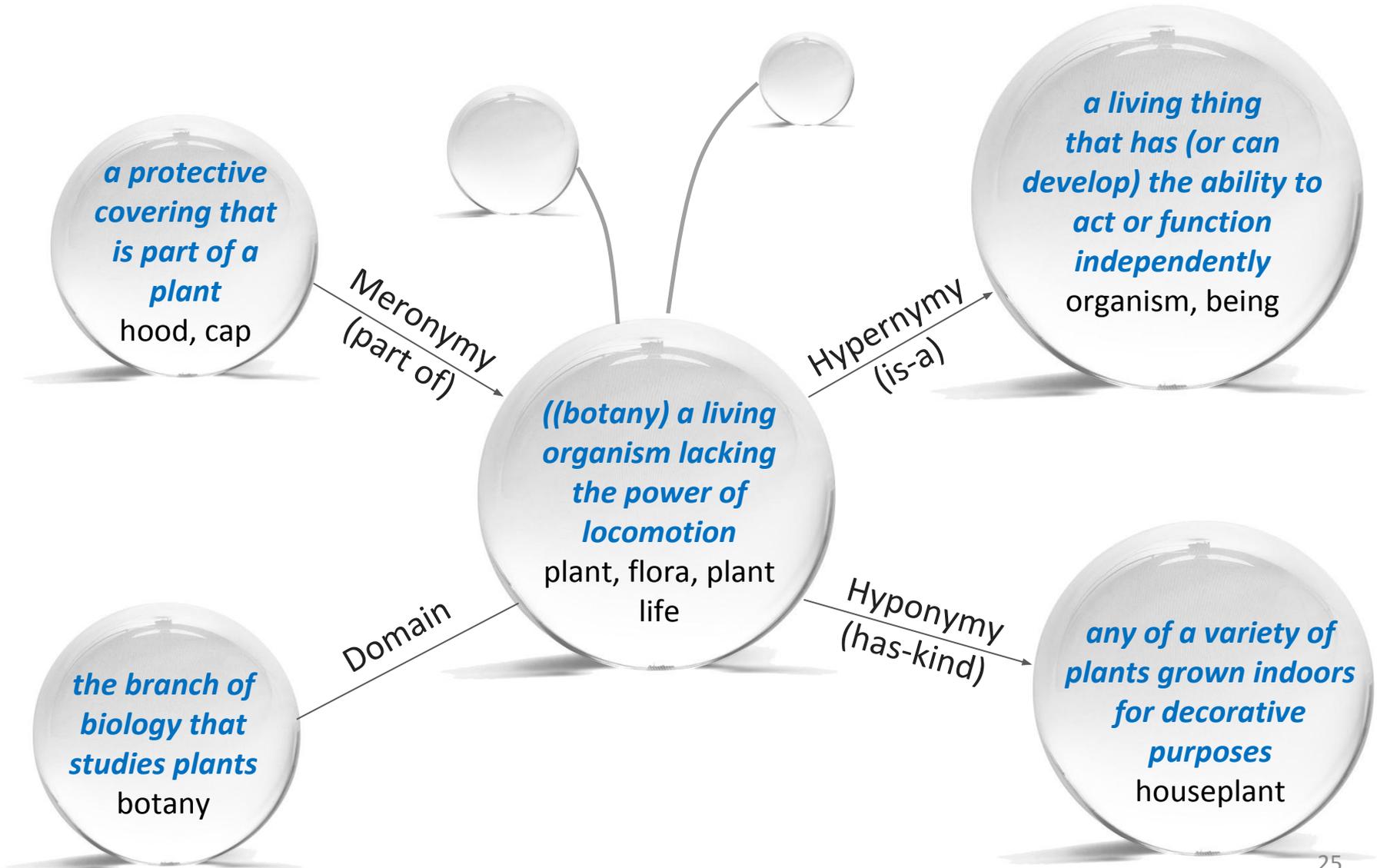


WordNet

Main unit: synset (concept)



WordNet semantic relations



Knowledge-based Representations (WordNet)

X. Chen, Z. Liu, M. Sun: **A Unified Model for Word Sense Representation and Disambiguation** (EMNLP 2014)

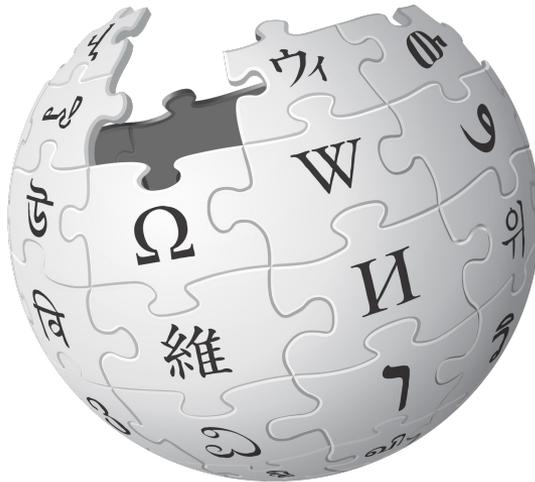
★ S. Rothe and H. Schutze: **AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes** (ACL 2015)

★ Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., & Smith, N. A. **Retrofitting Word Vectors to Semantic Lexicons** (NAACL 2015)*

S. K. Jauhar, C. Dyer, E. Hovy: **Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models** (NAACL 2015)

M. T. Pilehvar and N. Collier, **De-Conflated Semantic Representations** (EMNLP 2016)

Wikipedia



WIKIPEDIA
The Free Encyclopedia

Wikipedia

High coverage of **named entities** and **specialized concepts** from different domains

The screenshot displays the Wikipedia article for the University of California, Los Angeles (UCLA). The page layout includes a top navigation bar with 'Article' and 'Talk' tabs, and a search bar. The main content area features the article title 'University of California, Los Angeles' and a summary: 'From Wikipedia, the free encyclopedia'. The main text begins with a disambiguation note: '"UCLA", "Ucla", and "U.C.L.A." redirect here. For other uses, see UCLA (disambiguation)'. The article then describes UCLA as a public research university located in the Westwood neighborhood of Los Angeles, California, United States. It became the University of California Southern Branch in 1919, making it the second-oldest undergraduate campus of the ten-campus system after the original University of California campus in Berkeley (1873).^[11] It offers 337 undergraduate and graduate degree programs in a wide range of disciplines.^[12] With an approximate enrollment of 30,000 undergraduate and 12,000 graduate students, UCLA has the highest enrollment of any university in California^[6] and is the most applied to university in the United States with over 112,000 applications for fall 2015.^[13]

The university is organized into five undergraduate colleges, seven professional schools, and four professional health science schools. The undergraduate colleges are the College of Letters and Science; Henry Samueli School of Engineering and Applied Science (HSSEAS); School of the Arts and Architecture; School of Theater, Film, and Television; and School of Nursing. Fifteen^[14]^[15] Nobel laureates, one Fields Medalist,^[16] and three Turing Award winners^[17] have been faculty, researchers, or alumni. Among the current faculty members, 55 have been elected to the National Academy of Sciences, 28 to the National Academy of Engineering, 39 to the Institute of Medicine, and 124 to the American Academy of Arts and Sciences.^[18] The university was elected to the Association of American Universities in 1974.^[19]

UCLA student-athletes compete as the Bruins in the Pacific-12 Conference. The Bruins have won 125 national championships, including 112 NCAA team championships.^[20]^[21] UCLA student-athletes have won 250 Olympic medals: 125 gold, 65 silver and 60 bronze.^[22] The Bruins have competed in every Olympics since 1920 with one exception (1924), and have won a gold medal in every Olympics that the United States has participated in since 1932.^[23]

The sidebar on the left contains navigation links such as 'Main page', 'Contents', 'Featured content', 'Current events', 'Random article', 'Donate to Wikipedia', 'Wikipedia store', 'Interaction', 'Help', 'About Wikipedia', 'Community portal', 'Recent changes', 'Contact page', 'Tools', 'What links here', 'Related changes', 'Upload file', 'Special pages', 'Permanent link', 'Page information', 'Wikidata item', 'Cite this page', 'Print/export', and 'Create a book'. The sidebar on the right features the 'UCLA official seal' and a table of 'Former names' and 'Motto'.

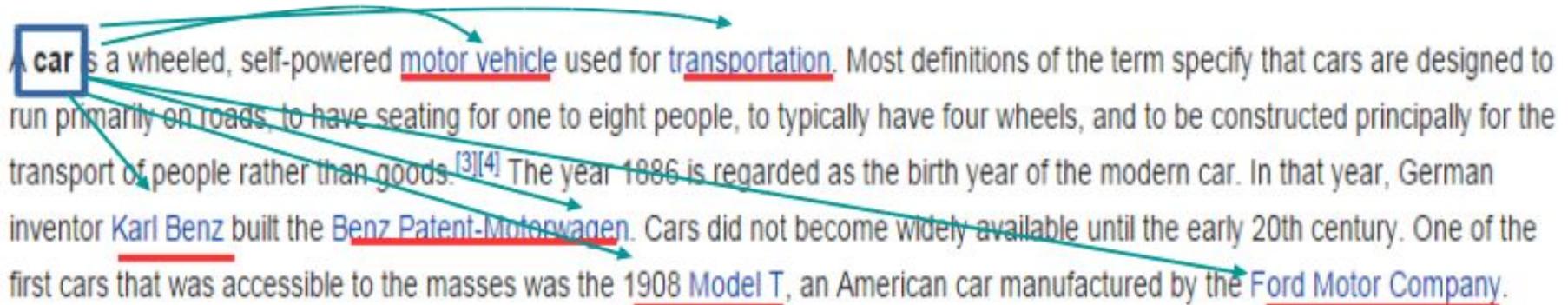
Former names	State Normal School at Los Angeles (1882-1919) University of California Southern Branch (1919-1927) University of California at Los Angeles (1927-1958)
Motto	Fiat lux (Latin)
Motto in English	Let there be light

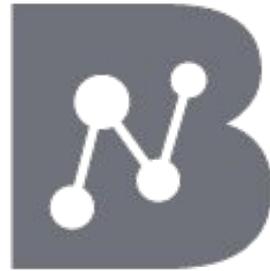
Wikipedia hyperlinks

A **car** is a wheeled, self-powered [motor vehicle](#) used for [transportation](#). Most definitions of the term specify that cars are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods.^{[3][4]} The year 1886 is regarded as the birth year of the modern car. In that year, German inventor [Karl Benz](#) built the [Benz Patent-Motorwagen](#). Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 [Model T](#), an American car manufactured by the [Ford Motor Company](#).

Wikipedia hyperlinks

A **car** is a wheeled, self-powered motor vehicle used for transportation. Most definitions of the term specify that cars are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods.^{[3][4]} The year 1886 is regarded as the birth year of the modern car. In that year, German inventor Karl Benz built the Benz Patent-Motorwagen. Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 Model T, an American car manufactured by the Ford Motor Company.

A diagram illustrating hyperlinks from the word "car" in the first sentence of the text. A blue box highlights the word "car". Five teal arrows originate from the box and point to the following underlined terms: "motor vehicle", "transportation", "Karl Benz", "Benz Patent-Motorwagen", and "Ford Motor Company".



BabelNet

Thanks to an automatic mapping algorithm, **BabelNet integrates Wikipedia and WordNet**, among other resources (Wiktionary, OmegaWiki, WikiData...).

Key feature: **Multilinguality** (271 languages)

BabelNet



BabelNet

ENTRA REGISTRATI

jaguar | ENGLISH | 4 SELEZIONATE | TRADUCI

PREFERENZE

Tutti | Concetti | Entità nominate | 21 risultati

Nome

Nome

Concept

Entity



jaguar, panther, Felis onca

A large spotted feline of tropical America similar to the leopard; in some classifications considered a member of the genus Felis

ID: 00033987n | Concetto

- ZH 美洲豹
- FR jaguar, panthère
- IT giaguaro, Panthera onca, pantera
- ES jaguar, panthera onca, pantera



Jaguar Cars, Jaguar

Jaguar Cars is a brand of Jaguar Land Rover, a British multinational car manufacturer headquartered in Whitley, Coventry, England, owned by Tata Motors since 2008.

ID: 00688731n | Entità

- ZH 捷豹
- FR Jaguar (automobile)
- IT Jaguar
- ES Jaguar Cars, Jaguar



Atari Jaguar, Jaguar (video game console)

The Atari Jaguar is a home video game console that was released by Atari Corporation in 1993.

ID: 02142312n | Entità

- ZH Atari Jaguar, 雅达利Jaguar
- FR Jaguar (console)
- IT Atari Jaguar
- ES Atari Jaguar



Mac OS X v10.2, Jaguar (macos)

Mac OS X version 10.2 Jaguar is the third major release of Mac OS X, Apple's desktop and server operating system.

- ZH Mac OS X Jaguar, Mac OS X v10.2
- FR Mac OS X v10.2

BabelNet

It follows the same structure of WordNet:
synsets are the main units

Nome



jaguar, panther, Felis onca

A large spotted feline of tropical America similar to the leopard; in some classifications considered a member of the genus Felis

ID: 00033987n | Concetto

ZH 美洲豹

FR jaguar, panthère

IT giaguaro, Panthera onca, pantera

ES jaguar, panthera onca, pantera

BabelNet

In this case, **synsets are multilingual**

Nome



jaguar, panther, Felis onca

A large spotted feline of tropical America similar to the leopard; in some classifications considered a member of the genus Felis

ID: [00033987n](#) | Concetto

- ZH 美洲豹
- FR jaguar, panthère
- IT giaguaro, Panthera onca, pantera
- ES jaguar, panthera onca, pantera

NASARI

(Camacho-Collados et al., AIJ 2016)

Goal

Build vector representations for multilingual BabelNet synsets.

How?

We exploit **Wikipedia semantic network** and **WordNet taxonomy** to construct a subcorpus (contextual information) for any given BabelNet synset.

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)

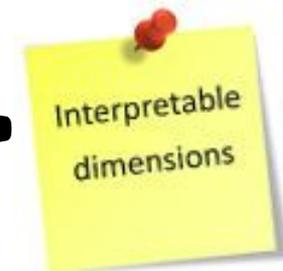
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded** (latent dimensions)

Three types of vector representations

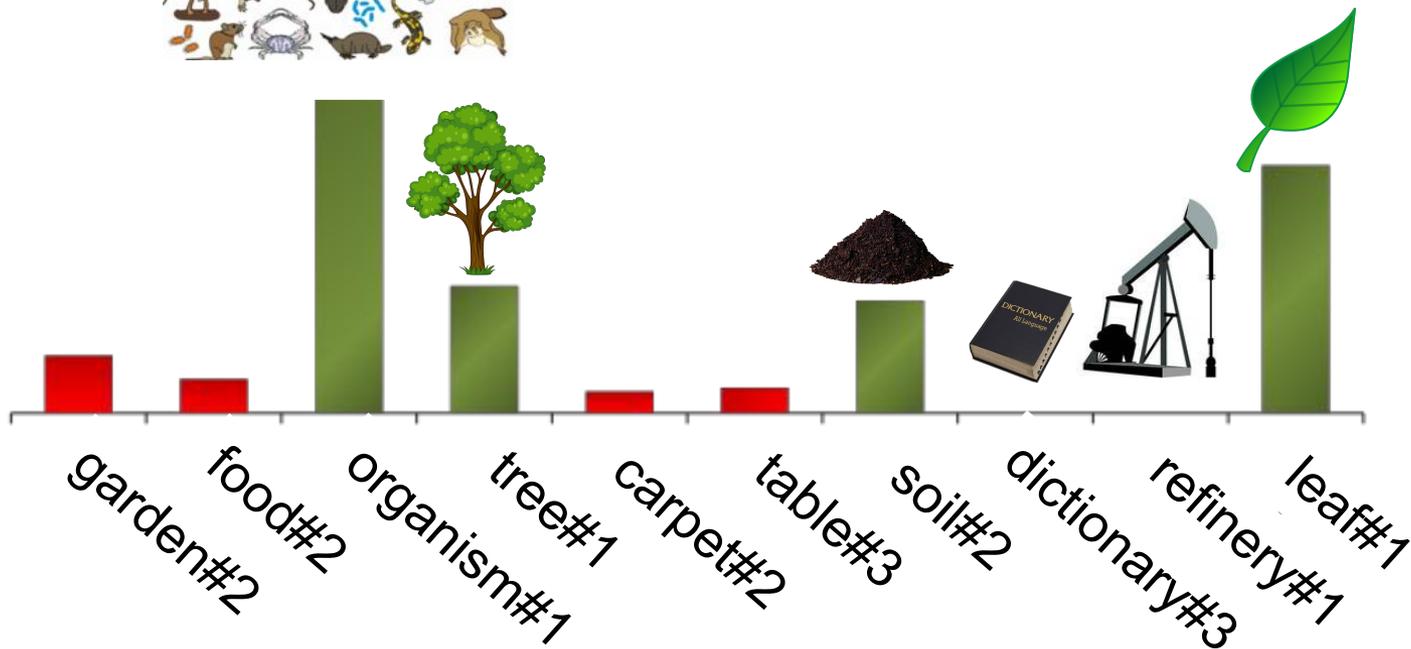
Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)
- **Embedded** (latent dimensions)



Human-interpretable dimensions

plant (living organism)



Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
 - **Unified** (dimensions are multilingual BabelNet synsets)
 - **Embedded**: Low-dimensional vectors exploiting **word embeddings** obtained from **text corpora**.
- 

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded:** Low-dimensional vectors exploiting **word embeddings** obtained from **text corpora**.

Word and synset embeddings share the same vector space!

Embedded vector representation

Closest senses



Bank (financial institution)

Closest senses	Cosine
Deposit account	0.99
Universal bank	0.99
British banking	0.98
German banking	0.98
Commercial bank	0.98
Banking in Israel	0.98
Financial institution	0.98
Community bank	0.97

Bank (geography)

Closest senses	Cosine
Stream bed	0.98
Current (stream)	0.97
River engineering	0.97
Braided river	0.97
Fluvial terrace	0.97
Bar (river morphology)	0.97
River	0.97
Perennial stream	0.96

bank

Closest senses	Cosine
Bank (financial institution)	0.86
Universal bank	0.86
British banking	0.86
German banking	0.85
Branch (banking)	0.85
McFadden Act	0.85
Four Northern Banks	0.84
State bank	0.84

SW2V

(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

SW2V

(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

How?

Updating the representation of the word and its associated senses interchangeably.

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses in context*.

*He withdrew money from the **bank**.*

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses* in context.

*He withdrew money from the **bank**.*



SW2V: Idea

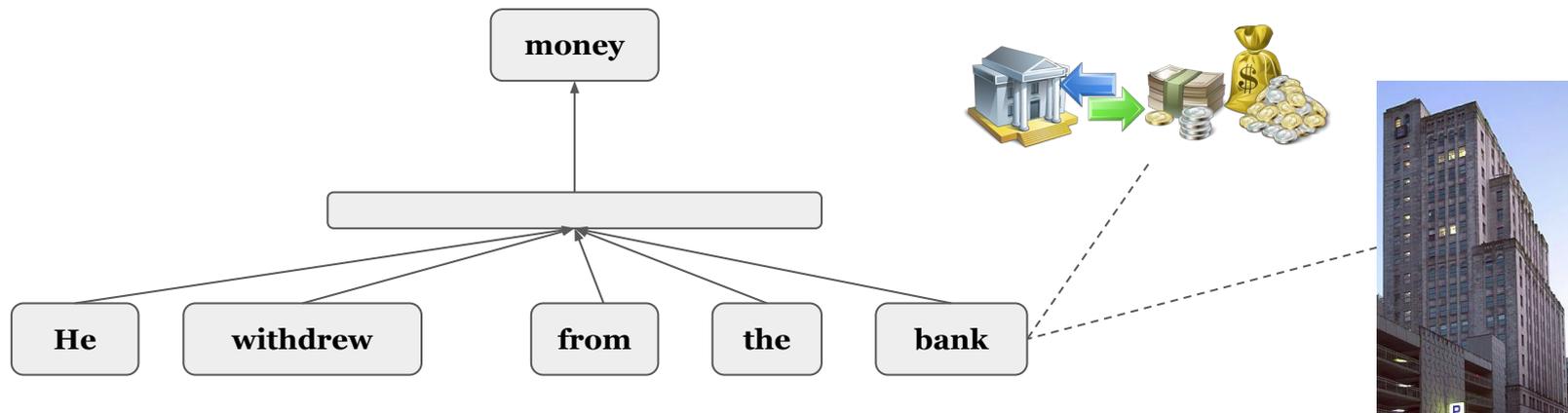
Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.
2. Use a **neural network** where the **update of word and sense embeddings is linked**, exploiting *virtual* connections.

SW2V: Idea

Given as input a corpus and a semantic network:

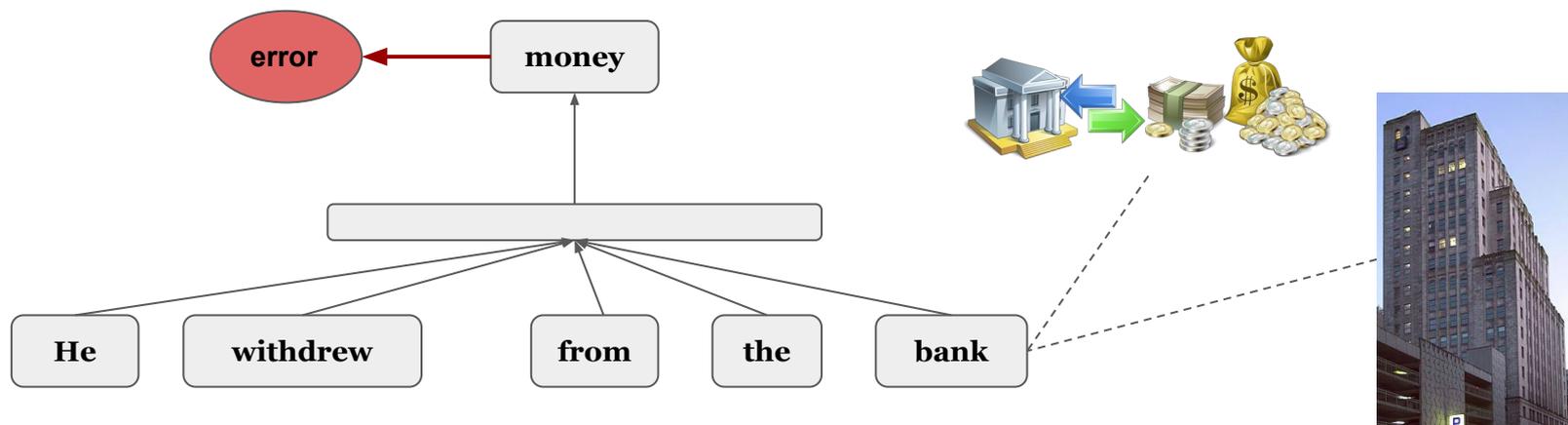
1. Use a semantic network to link to each word its *associated senses* in context.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

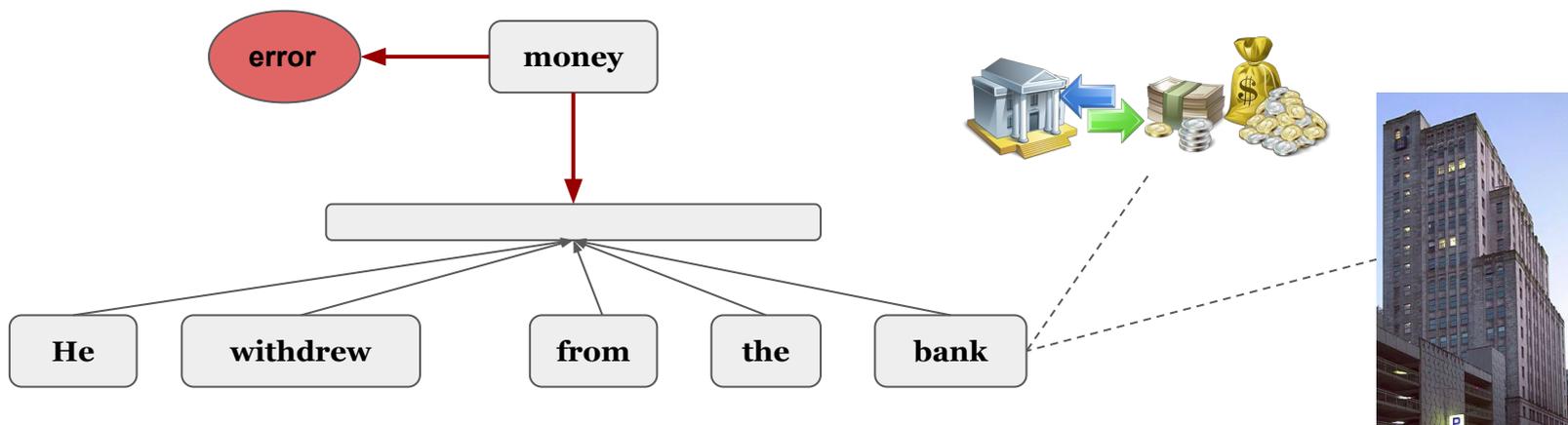
1. Use a semantic network to link to each word its *associated senses* in context.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

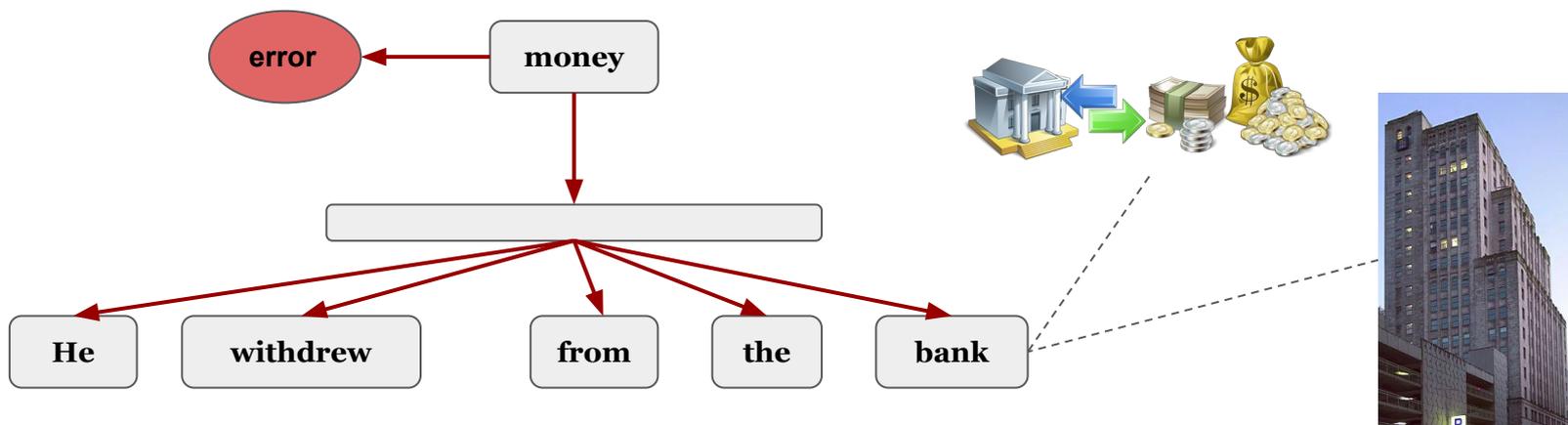
1. Use a semantic network to link to each word its *associated senses* in context.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

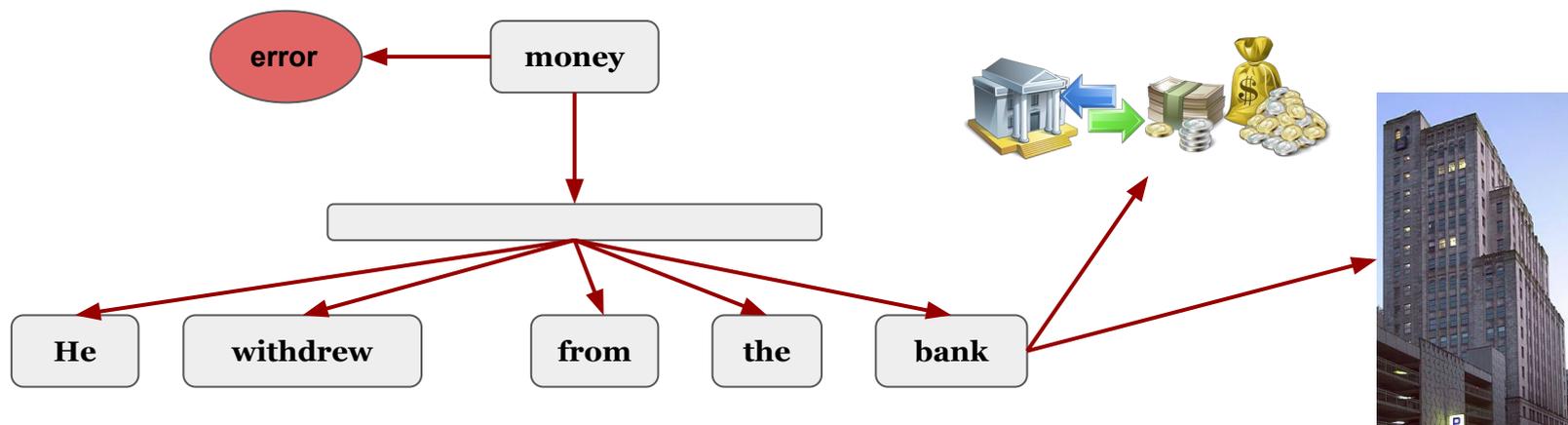
1. Use a semantic network to link to each word its *associated senses* in context.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses* in context.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

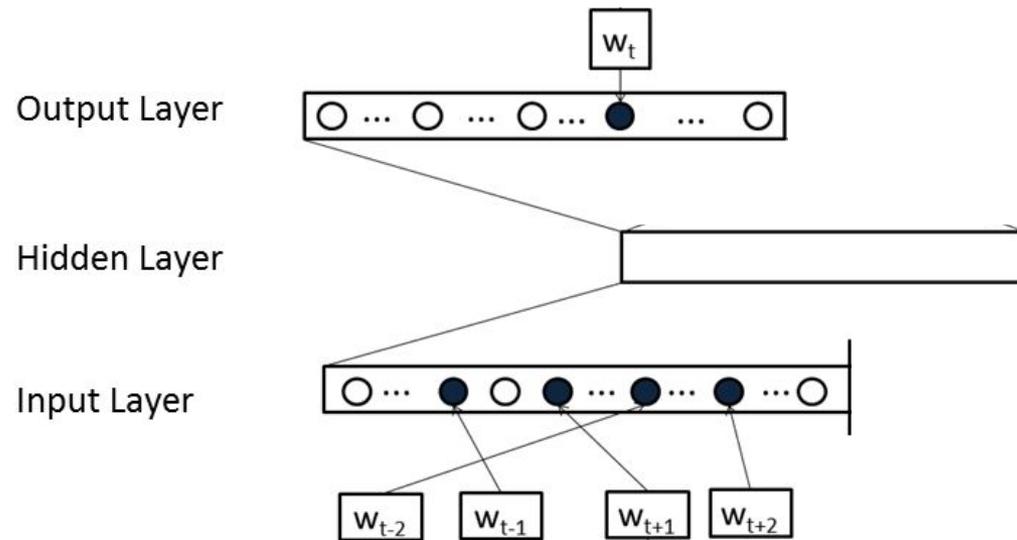
Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.
2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

*In this way it is possible to learn word and sense/synset embeddings jointly on a **single training**.*

Full architecture of W2V (Mikolov et al., 2013)

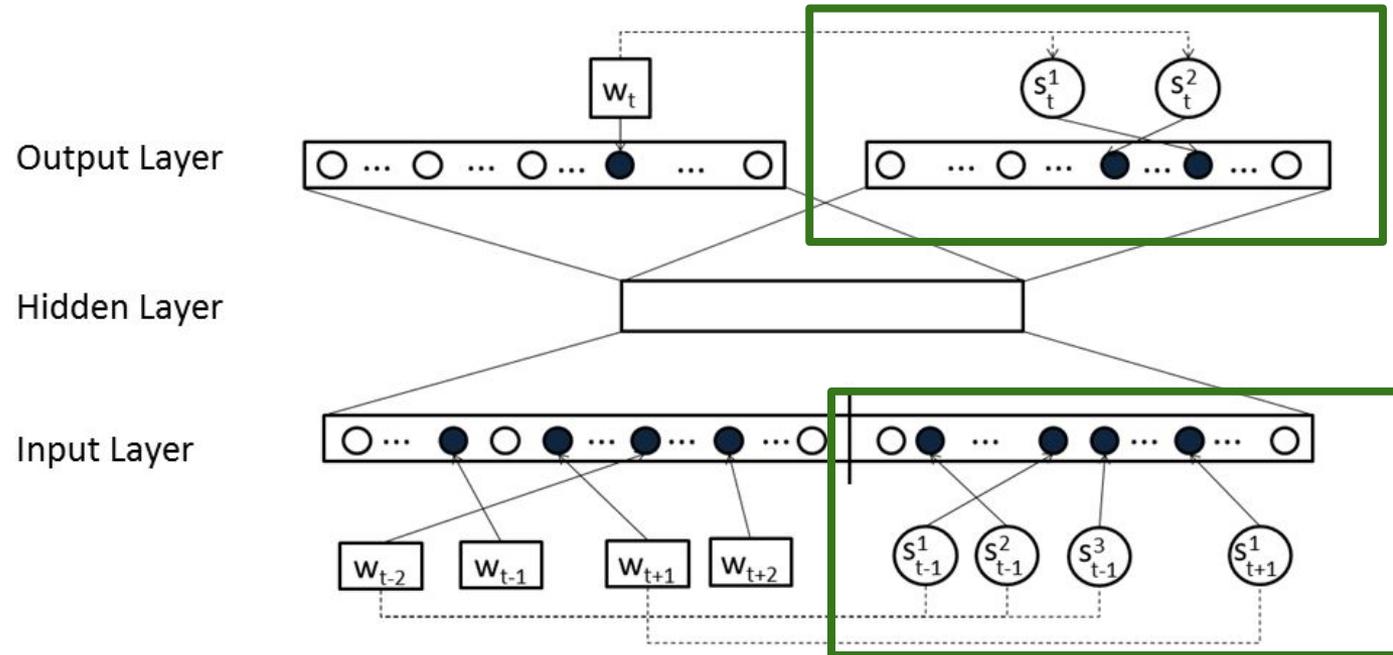
$$E = -\log(p(w_t | W^t))$$



Words and associated senses used both as input and output.

Full architecture of SW2V

$$E = -\log(p(w_t | W^t, \mathbf{S}^t)) - \sum_{s \in S^t} \log(p(s | W^t, \mathbf{S}^t))$$



Words and associated senses used both as input and output.

Word and senses connectivity: example 1



*company*_n² (*military unit*)

AutoExtend

company_n⁹

company

company_n⁸

company_n⁶

company_n⁷

company_v¹

firm

business_n¹

firm_n²

company_n¹

SW2V

battalion_n¹

battalion

regiment_n¹

detachment_n⁴

platoon_n¹

brigade_n¹

regiment

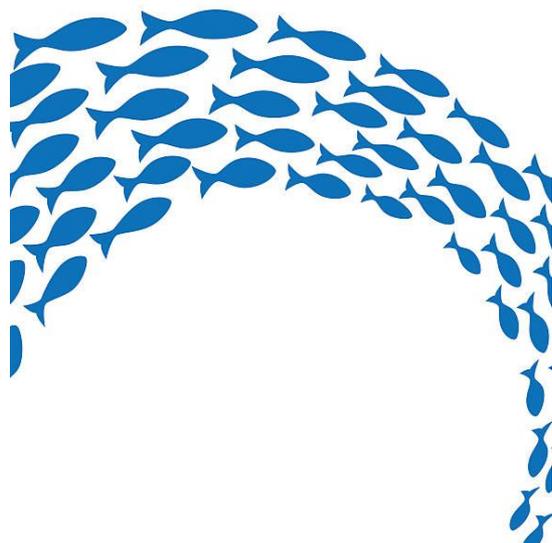
corps_n¹

brigade

platoon

**Ten closest word and sense embeddings
to the sense *company* (*military unit*)**

Word and senses connectivity: example 2



school_n⁷ (group of fish)

AutoExtend

school
school_n⁴
school_n⁶
school_v¹
school_n³
elementary
schools
elementary_a³
school_n⁵
elementary_a¹

SW2V

schools_n⁷
sharks_n¹
sharks
shoals_n³
fish_n¹
dolphins_n¹
pods_n³
eels
dolphins
whales_n²

**Ten closest word and sense embeddings
to the sense *school (group of fish)***

Applications of knowledge-based sense representations

- **Taxonomy Learning** (Espinosa-Anke et al., EMNLP 2016)
- **Open Information Extraction** (Delli Bovi et al. EMNLP 2015).
- **Lexical entailment** (Nickel & Kiela, NIPS 2017)
- **Word Sense Disambiguation** (Rothe & Schütze, ACL 2015)
- **Sentiment analysis** (Flekova & Gurevych, ACL 2016)
- **Lexical substitution** (Cocos et al., SENSE 2017)
- **Computer vision** (Young et al. ICRA 2017)

...

Applications

- ❖ Domain labeling/adaptation
- ❖ Word Sense Disambiguation
- ❖ Downstream NLP applications (e.g. text classification)

Domain labeling

(Camacho-Collados and Navigli, EACL 2017)

Annotate each **concept/entity** with its corresponding **domain of knowledge**.

To this end, we use the [Wikipedia featured articles page](#), which includes 34 domains and a number of Wikipedia pages associated with each domain (*Biology, Geography, Mathematics, Music, etc.*).

Domain labeling

How to associate a concept with a domain?

1. Learn a **NASARI vector** for the concatenation of all Wikipedia pages associated with a given domain.
2. Exploit the **semantic similarity** between knowledge-based vectors and **graph properties** of the lexical resources.

BabelDomains

(Camacho-Collados and Navigli, EACL 2017)

As a result:

Unified resource with information about domains of knowledge

*BabelDomains available for **BabelNet**, **Wikipedia** and **WordNet** available at*

<http://icl.uniroma1.it/babeldomains>

Already integrated into BabelNet (online interface and API)

BabelDomains



BabelNet

LOG IN REGISTER

eclipse ENGLISH TRANSLATE INTO... SEARCH

PREFERENCES

All Concepts Named Entities 48 results

🎵 🖱️ 🌐 ⭐ 🎮 ⚽ 📊 📄 🖼️ 🌍 🎬 📈

- Noun
- Verb

Noun

Physics and astronomy



eclipse, occultation

One celestial body obscures another

ID: 00029648n | Concept

Computing



Eclipse (software)

In computer programming, Eclipse is an integrated development environment.

ID: 01457115n | Named Entity

Media



The Twilight Saga: Eclipse, Eclipse (2010 film)

The Twilight Saga: Eclipse, commonly referred to as Eclipse, is a 2010 American romantic fantasy film based on Stephenie Meyer's 2007 novel Eclipse.

ID: 01455414n | Named Entity

Domain filtering for supervised distributional hypernym discovery

(Espinosa-Anke et al., EMNLP 2016;
Camacho-Collados and Navigli, EACL 2017)



Apple

is | a

Fruit

Task: Given a term, predict its hypernym(s)

Model: Distributional supervised system based on the transformation matrix of Mikolov et al. (2013).

Idea: Training data filtered by domain of knowledge

Domain filtering for supervised distributional hypernym discovery

Domain-filtered training data

	art			biology			education			geography			health		
Train	MRR	MAP	R-P												
5k	0.12	0.12	0.12	0.63	0.63	0.59	0.00	0.00	0.00	0.08	0.07	0.07	0.08	0.08	0.07
15k	0.21	0.20	0.18	0.84	0.72	0.79	0.22	0.22	0.21	0.15	0.14	0.14	0.08	0.07	0.07
25k	0.29	0.27	0.26	0.84	0.83	0.81	0.33	0.32	0.30	0.23	0.22	0.21	0.09	0.09	0.08
25k+ K_{1k}^d	0.29	0.28	0.26	0.84	0.80	0.79	0.32	0.29	0.27	0.22	0.22	0.21	0.09	0.09	0.08
25k+ K_{25k}^d	0.26	0.24	0.22	0.70	0.63	0.56	0.38	0.36	0.33	0.15	0.13	0.12	0.11	0.11	0.10
25k+ K_{50k}^r	0.28	0.26	0.24	0.82	0.77	0.72	0.36	0.33	0.30	0.17	0.16	0.16	0.12	0.11	0.10
100k $_{wd}^r$	0.00	0.00	0.00	0.84	0.81	0.77	0.00	0.00	0.00	0.01	0.01	0.01	0.07	0.06	0.06
100k $_{kbu}^r$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.12	0.12	0.11
Baseline	0.13	0.12	0.10	0.58	0.57	0.57	0.10	0.10	0.09	0.12	0.09	0.05	0.07	0.13	0.14

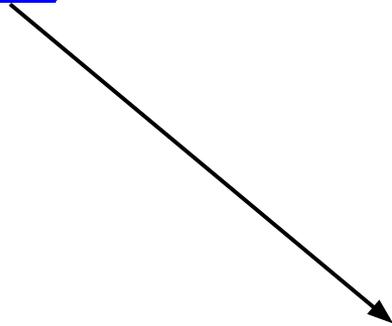
Non-filtered training data

Results on the hypernym discovery task for five domains

Conclusion: Filtering training data by domains prove to be clearly beneficial

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]



Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

X



Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]



Word Sense Disambiguation

(Camacho-Collados et al., AIJ 2016)

Basic idea

Select the sense which is semantically closer to the semantic representation of the whole document
(global context).

$$\hat{d}(s) = \operatorname{argmax}_{d \in D} WO(\vec{N}_{ASARI_{lex}}(s), \vec{v}_{lex}(d))$$

Word Sense Disambiguation on textual definitions

(Camacho-Collados et al., LREC 2016; LREV 2018)

Combination of a graph-based disambiguation system (Babelfy) with NASARI to **disambiguate** the concepts and named entities of **over 35M definitions** in **256 languages**.

Sense-annotated corpus freely available at

<http://lcl.uniroma1.it/disambiguated-glosses/>

Context-rich WSD



castling (*chess*)



*Interchanging the positions of the **king** and a **rook**.*

Context-rich WSD



castling (*chess*)



*Interchanging the positions of the **king** and a **rook**.*



***Castling** is a move in the game of **chess** involving a player's **king** and either of the player's original **rooks**.*



*A move in which the **king** moves two **squares** towards a **rook**, and the **rook** moves to the other side of the **king**.*

Context-rich WSD



castling (*chess*)



*Interchanging the positions of the **king** and a **rook**.*



Castling is a move in the game of **chess** involving a player's **king** and either of the player's original **rooks**.



A move in which the **king** moves two **squares** towards a **rook**, and the **rook** moves to the other side of the **king**.



Manœuvre du jeu
d'échecs



Rošáda je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.



Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden



El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.



Rokade er et spesialtrekk i **sjakk**.



Rok İngilizce'de kaleye **rook** denmektedir.



Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.

Context-rich WSD



castling (chess)



Interchanging the positions of the **king** and a **rook**.



Castling is a move in the game of **chess** involving a player's **king** and either of the player's **rooks**.



which the **king** moves two squares towards a **rook**, and the **rook** moves to the other side of the **king**.



Mano d'échec



Rošáda je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.



Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden



El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.



Rokade er et spesialtrekk i **sjakk**.



Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.



Rok İngilizce'de kaleye **rook** denmektedir.

Context-rich WSD exploiting parallel corpora

(Delli Bovi et al., ACL 2017)

Applying the same method to provide **high-quality sense annotations from parallel corpora** (Europarl): 120M+ sense annotations for 21 languages.

<http://lcl.uniroma1.it/eurosense/>

Extrinsic evaluation: Improved performance of a standard supervised WSD system using this automatically sense-annotated corpora.

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect -> **Solution:** High-confidence disambiguation

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect -> **Solution:** High-confidence disambiguation
- WordNet lacks coverage

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect -> **Solution:** High-confidence disambiguation
- WordNet lacks coverage -> **Solution:** Use of Wikipedia

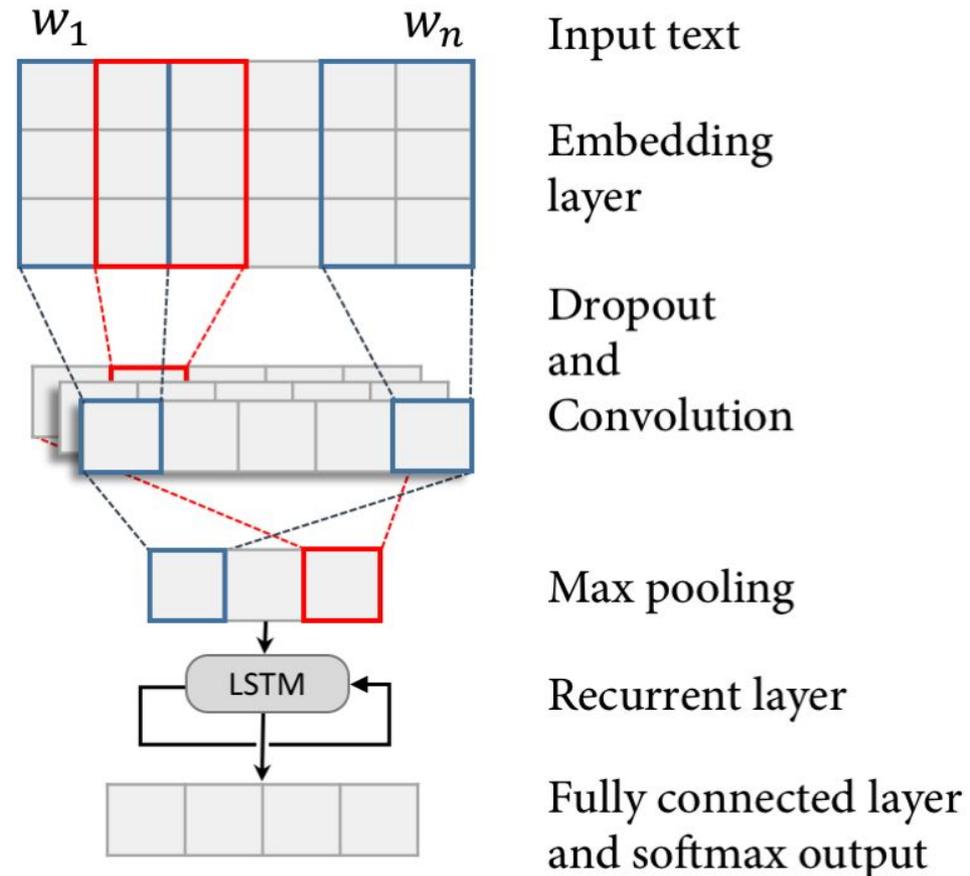
Tasks: Topic categorization and sentiment analysis (polarity detection)

Topic categorization: Given a text, assign it a topic (e.g. politics, sports, etc.).

Polarity detection: Predict the sentiment of the sentence/review as either positive or negative.

Classification model

Standard CNN classifier
inspired by Kim (2014)

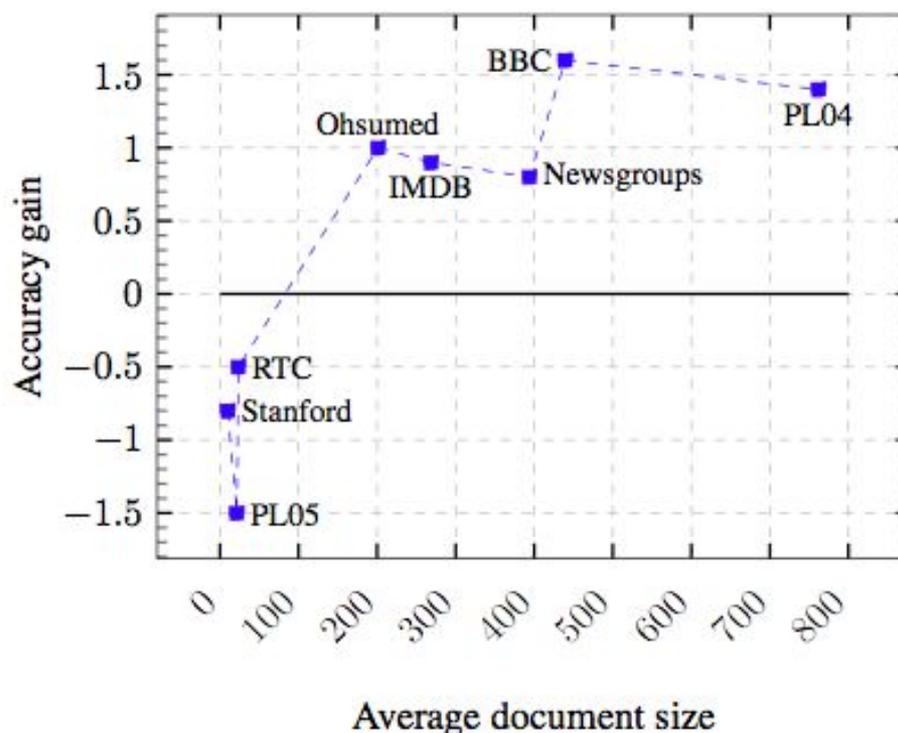


Sense-based vs. word-based: Conclusions

Sense-based **better** than word-based... when
the **input text is large enough**

Sense-based vs. word-based:

Sense-based **better** than word-based... when the **input text is large enough**:



Why does the input text size matter?

- Word sense disambiguation works better in larger texts (Moro et al. 2014; Raganato et al. 2017)
- Disambiguation increases sparsity

Contextualized word embeddings ELMo/BERT



Peters et al.
(NAACL 2018)



Devlin et al.
(NAACL 2019)

Contextualized word embeddings

ELMo/BERT



New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse



Contextualized word embeddings ELMo/BERT



As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

Contextualized word embeddings

ELMo/BERT



As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Contextualized word embeddings

ELMo/BERT



As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Important **improvements in many NLP tasks**.

Contextualized word embeddings

ELMo/BERT (examples)



*He withdrew money from the **bank**.*

*The **bank** remained closed yesterday.*

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings ELMo/BERT (examples)



0.25, 0.32, -0.1

*He withdrew money from the **bank**.*

0.22, 0.30, -0.08

*The **bank** remained closed yesterday.*

-0.8, 0.01, 0.3

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings

ELMo/BERT (examples)



Similar vectors

0.25, 0.32, -0.1

*He withdrew money from the **bank**.*

0.22, 0.30, -0.08

*The **bank** remained closed yesterday.*

-0.8, 0.01, 0.3

*We found a nice spot by the **bank** of the river.*

How well do these models capture “meaning”?



Good enough for many applications.

Room for improvement. No noticeable improvements in:

- *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%
- *Word-in-Context Challenge*: BERT ~65% **vs** Humans ~85%

How well do these models capture “meaning”?



Good enough for many applications.

Room for improvement. No noticeable improvements in:

- *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%

→ *requires commonsense reasoning*

- *Word-in-Context Challenge*: BERT ~65% **vs** Humans ~85%

→ *requires abstracting the notion of sense*

For more information on meaning representations (embeddings):

- ❖ ACL 2016 Tutorial on “**Semantic representations of word senses and concepts**”:
http://josecamachocollados.com/slides/Slides_ACL16Tutorial_SemanticRepresentation.pdf
- ❖ EACL 2017 workshop on “**Sense, Concept and Entity Representations and their Applications**”: <https://sites.google.com/site/senseworkshop2017/>
- ❖ NAACL 2018 Tutorial on “**Interplay between lexical resources and NLP**”:
<https://bitbucket.org/luisespinoza/lr-nlp/>
- ❖ “**From Word to Sense Embeddings: A Survey on Vector Representations of Meaning**” (JAIR 2018): <https://www.jair.org/index.php/jair/article/view/11259>

Thank you!

Questions please!

camachocolladosj@cardiff.ac.uk

[@CamachoCollados](#)

josecamachocollados.com